INTERNATIONAL STANDARD

ISO/IEC 20382-1

First edition
2017-10

Information technology — User interfaces — Face-to-face speech translation —

Part 1: User interface

Technologies de l'information — Interface utilisateur — Face-à-face discours traduction —
Partie 1: Interface utilisateur

Reference number
ISO/IEC 20382-1:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form a specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organizations to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

A list of all parts in the ISO/IEC 20382 series can be found on the ISO website.

Introduction

It is important to consider people with special requirements to ensure that they can gain the same benefits from ICT. One of those special requirements is to help people to avoid language barriers in global environments. Automatic speech translation systems have existed for a long time, but they have functional limitations as well as technical ones with regard to usability and accessibility.

One reason for these limitations is the diversity of the languages currently used. It is difficult to support many languages by one or several speech translation systems. A flexible and interoperable standardized framework is needed to work with all different languages, utilizing many speech translation systems already developed in many countries. Other considerations to make a natural and usable speech translation service possible include applying users' characteristics within the system, such as emotion, speech style, gender type and other attributes. To reflect those characteristics in the output speech translation, a standardized user interface should reflect the input and output data and transfer them to the user's device.

The main purpose of this document is to help users of different languages by providing speech translation service in easier and more convenient ways with a standardized user interface for face-to-face speech translation.

Information technology — User interfaces — Face-to-face speech translation —

Part 1:
User interface

1 Scope

This document specifies face-to-face speech translation designed to interoperate among multiple translation systems with different languages. It also specifies the speech translation features, general requirements and functionality, thus providing a framework to support a convenient speech translation service in face-to-face situations. This document is applicable to user interfaces for speech translation and communication protocols for setting up a translation session among users. This document is not applicable to defining the speech translation engine itself.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 20382-2:2017, *Information technology — User interface — Face-to-face speech translation — Part 2: System architecture and functional components*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

The ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/;

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
face-to-face
arrangement where two users are physically in the same location

3.2
short range wireless communication
SRWC
wireless transmission that uses signals that travel from a few centimetres to several metres

EXAMPLE Bluetooth.

4 Abbreviated terms

AI audio indicator

F2F face-to-face

TTS text to speech

VI video indicator

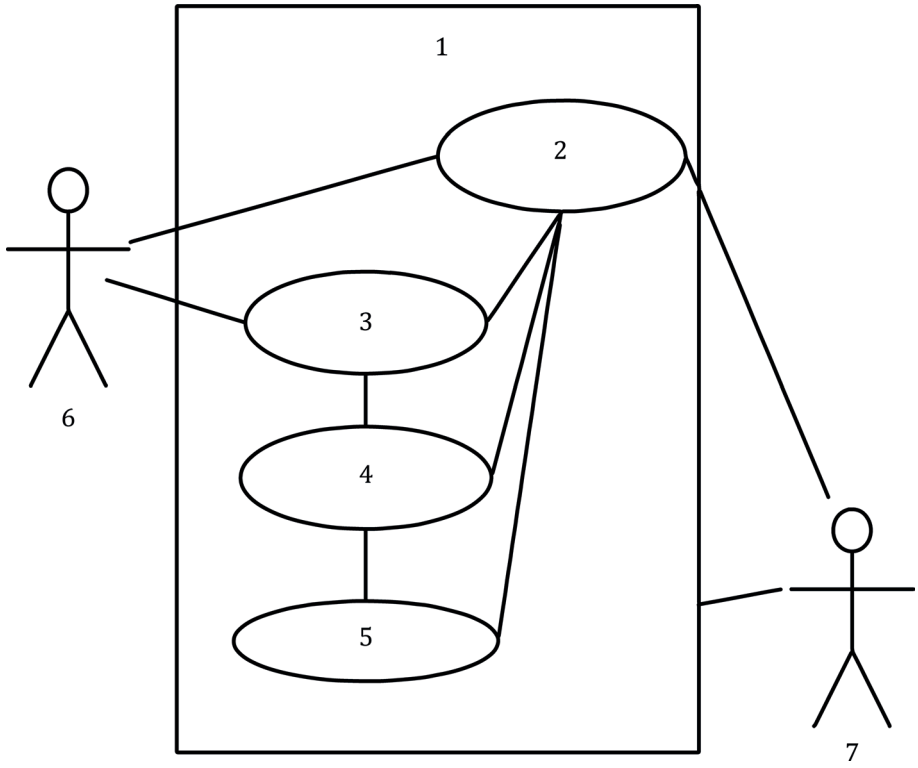
WD wearable device

MD mobile device

5 Overview of face-to-face speech translation

5.1 General

A face-to-face speech translation system enables users of different languages to communicate with each other with spoken languages in a face-to-face situation by providing machine translation results (as in Figure 1). (For standardization activities for speech translation, see Annex A). In a face-to-face speech translation system, mobile devices and wearable devices, such as earphones, are used for convenient user experiences. The main functions of wearable devices in the translation system are the processing of input and output of speech signals as a microphone and speakers. Speech recognition and speech synthesis are performed in each user's mobile device. The machine translation function resides in the translation servers.



Key

1 F2F speech translation system
2 UI set–up
3 speech recognition (see ISO/IEC 20382-2)
4 language translation (see ISO/IEC 20382-2)
5 speech synthesizer (see ISO/IEC 20382-2)
6 speaker
7 listener

Figure 1 — Service example of face-to-face speech translation

5.2 Service flow

The following steps are typical speech translation service processes in a face-to-face speech translation system. For more information, see ISO 20382-2:2017, 6.2 and Figure 2.

1) After a session connection is made between user A and user B, speech is input by user A in language A. The speech signal is transmitted to the mobile device of user A.

2) The speech is recognized by a speech recognition module in the mobile device of user A. Then, the translation operation is requested to machine translation server K.

3) The translation is performed at server K and the result, which is in language B, is sent back to the mobile device of user A in text form.

4) The translation result is then sent to the mobile device of user B through SRWC communication.

5) The translation result, which is in text form, is transformed to a speech signal in language B by a speech synthesizer (TTS) and sent to the wearable device of user B.

6) For user B's speech, steps (6) to (10) are performed in the same way and the session ends.

5.3 Service types

There are two service types depending on the number of participants in the translation session:

— two way translation: Two users are participating in the translation session;

— multi-way translation: More than three users are participating in the translation session. The participants may start the dialogue at different points in time.

5.4 Service mode

Several service modes are selected for usability in different situations.

— Open mode: In this mode, the dialogue is not protected and can be heard by the public. Any user can barge into the dialogue.

— Protected mode: In this mode, only allowed users can participate in the translation session and privacy is guaranteed. The dialogue is protected and not heard by the public.

— Automatic mode: The session starts and ends by automatic operation.

— Manual mode: In this mode, the user can decide to be connected to other users manually.

5.5 Service situation

Several situations are classified as crowded and non-crowded depending on the number of candidates of users.

— Crowded situation: In this situation, many candidates are nearby for the translation session.

— Non-crowded situation: In this situation, only one or two candidates are nearby for the translation session.

6 Functional requirements

6.1 User communication requirements

6.1.1 General

6.1.2 Required

— The translation system shall allow the users to start a translation session with less than 3 operations.

NOTE The user is able to start a translation session with as few operations as possible.

— The translation system shall allow the users to start a translation session within 10 seconds.

6.1.3 Optional

— The translation system should allow the users to have a session with multiple users.

— The translation system should allow the users to have additional participants after the session starts.

— The translation system should allow the users to have a session with available target users by approaching them.

6.2 User interface requirements

6.2.1 General

6.2.1.1 Required

— There shall be no restriction on the users to use the translation system.

NOTE Any user can use the user interface of the translation system.

6.2.1.2 Optional

— The results of speech recognition and translation should be displayed.

— The user should be able to edit the text output of the speech recognition module.

— Frequently used functions should be shown on the top-level menu.

— The depth of menu should not exceed 4 steps.

— There should be a button to go back to the top-level menu.

— The user interface should provide options to select alternate translation results.

NOTE The user chooses among the translation results, speech or text, on the screen.

— The user interface should provide options for preferences such as gender, emotion, speech style and other features defined in the user profile.

— The user interface should provide the functionality that reflects the user's characteristics defined in the user profile.

6.3 User device requirements

6.3.1 General

6.3.1.1 Required

None.

6.3.1.2 Optional

— The translation device should provide a function to allow the user to show the signal to participate in the current translation session.

— The translation devices should operate easily.

NOTE If the time to start operating a specific function takes more than 3 minutes, it does not satisfy this requirement.

6.4 Accessibility requirements

6.4.1 General

6.4.1.1 Required

— The input from the user shall be in a text form as well as a speech form for people with speaking disabilities.